The *GLAST* Science Support Center and the High Energy Astrophysics Science Archive Research Center Interface Control Document

Michael F. Corcoran
Code 662,
NASA/GSFC/USRA,
Greenbelt,
MD 20771

David S. Davis
Code 661,
NASA/GSFC/UMBC,
Greenbelt,
MD 20771

Donald J. Horner
Code 660.1
NASA/GSFC/Wyle Labs
Greenbelt,
MD 20771

SUMMARY

This document details the manner in which *GLAST* archival data (event files, lightcurves etc) and metadata (source catalogs, as-flown timelines and others) and other mission assets are transferred from the *GLAST* Science Support Center to the *HEASARC*.

LOG OF SIGNIFICANT CHANGES

Release Date	**Sections Changed**	**Brief Notes**
2006 Apr 28	All	First (internal) Draft
2006 Sep 26	§§7, 10.2	added discussion of web-based "tools"; re-organized description of metadata to come after discussion of data
2007 Feb 08	§6.8	updated Notification and Verification subsections
2007 Sep 11	§6.4.1	added
2007 Sep 13	all	after discussion with D. Horner
2008 Feb 11	§6.7 &Figure 1	after discussion with P. Newman, D. Horner, D. Davis
2008 Feb 20	§§6,6.5, 6.5, 6.9, 8, Table 6.7 & Figures 1, 3, **??**, 5, 6	after discussion with D. Horner & E. Sabol
2008 Mar 31	§§6,6.5, 6.5, 6.9, 8, Table 6.7	after discussion with D. Horner & E. Sabol

Contents

List of Tables

1 Purpose and Scope

As described in the *GLAST* Science Support Center (*GSSC*) Functional Requirements Document (Band, D., 2006b), the GSSC is responsible for supporting the astronomical community's use of *GLAST* data by disseminating *GLAST* data and results, by providing analysis software and expertise, and by archiving *GLAST* data. At the end of the *GLAST* mission and the close of the *GSSC*, the responsibility for maintaining the *GLAST* archive devolves onto the High Energy Astrophysics Science Archive Research Center (*HEASARC*). This Interface Control Document describes the manner in which data from the *GLAST* mission and other mission assets (software, hardware, documentation) which are managed by the *GSSC* are transferred to and archived at the High Energy Astrophysics Science Archive Research Center (*HEASARC*). This document describes the types of *GLAST* data and how these data will be archived and accessed and how data, analysis software, user guides and other documentation, and other mission assets will be supported at the *HEASARC* after the end of the *GLAST* mission and the close of the *GSSC*. It describes the roles, responsibilities, and relationship of the *GSSC* and the *HEASARC* in regards to safeguarding *GLAST* mission assets.

2 Relevant Documents

- GLAST Project Management Data Plan (433-PLAN-009) (Band, D., 2006a)
- GSSC-Functional Requirements Document (Rev. A) (433-RQMT-0002) (Band, D., 2006b)
- *GSSC-HEASARC* Memorandum of Understanding (GSSC-0008) (Davis, D. S. and Corcoran, M. F., 2006)
- Ingest System Detailed Design Document (GSSC-0009)(Horner, D., 2006)
- Science Data Products File Format Document (Band, D., 2006c)
- Science Data Products Interface Control Document (GLAST-GS-ICD-0006) (Band, D., 2006d)

3 Mission Overview

The Gamma-Ray Large Area Space Telescope (*GLAST*) mission is scheduled for launch no earlier than May 2008. The *GLAST* observatory includes 2 instruments, the Large Area Telescope (LAT) and the *GLAST* Burst Monitor (GBM). A brief description of these two instruments, largely derived from the Project Data Management Plan (Band, D., 2006a), is given below.

3.1 The Large Area Telescope

As discussed in the *GLAST* Project Data Management Plan, the LAT's principal objective is high sensitivity gamma-ray observations of celestial sources in the energy range from $\sim$ 20 MeV to $>$ 300 GeV. The LAT is composed of an array of 16 tracker (TKR) modules, 16 calorimeter (CAL) modules, and a segmented anticoincidence detector (ACD). The LAT will have a wide FOV (> 2 sr), large effective area ($> 8,000$ cm^2 maximum effective area at normal incidence at a few GeV), and excellent angular resolution (on-axis single photon 68% space containment angle: $< 0.15^\circ$ for $E >$10 GeV and $< 3.5^\circ$ for $E = 100$ MeV). The LAT will provide good energy resolution ($\Delta E/E < 10\%$ in the central part of the energy range) to enable spectral studies of high-energy sources. The actual values may be better than the required values given here. The LAT will detect point sources that are more than 200 times fainter than the Crab nebula. For strong point sources, the position will be determined to about 0.5 arcminute. Spectra will be measurable over the entire energy range for the stronger sources. The large FOV and low dead time ($< 100\mu$s/event required) of the LAT will allow it to monitor the sky for high-energy transients, particularly Gamma-Ray Bursts (GRBs). On-board data processing will allow near real-time notification to the ground of transients. The Principal Investigator of the LAT is Prof. Peter Michelson (Stanford). The GLAST LAT Collaboration includes scientists from the following universities and scientific organizations:

Stanford University, including the Stanford Linear Accelerator (SLAC); Goddard Space Flight Center; University of California at Santa Cruz; Naval Research Laboratory (NRL); University of Washington; Sonoma State University; Texas A&M University-Kingsville; Stockholm University and Royal Institute of Technology, Stockholm; Commissariat a lEnergie Atomique, Departement dAstrophysique, Saclay, France; Institut National de Physique Nuclearie et de Physique des Particules, France; Instituto Nazionale di Fisica Nucleare, Italy; Agenzia Spaziale Italiana, Italy; Instituto di Fisica Cosmica, CNR; Hiroshima University; Institute of Space and Astronautical Science, Tokyo; Riken; Tokyo Institute of Technology.

In addition, the LAT team includes Affiliated Scientists from 29 institutions world-wide.

3.2 The *GLAST* Burst Monitor

The GBM will provide simultaneous spectral and temporal measurements in the energy range 5 keV – 30 MeV for all GRBs within the LAT FOV. The GBM will have a timing accuracy of $\sim 2\mu$s and an energy resolution of about 10% in the 0.1–1 MeV range. The average GBM statistical location uncertainty for all detected GRBs is estimated to be $\sim 15^\circ$ ($1-\sigma$ radius), improving to $\sim 9^\circ$ ($\sim 1.5^\circ$) for the brightest 40% (5%) of the bursts. The systematic location error is estimated to be $\sim 5-10^\circ$ for on-board processing and $\sim 1-2^\circ$ for final ground processed data. The combined GBM and LAT effective energy range will span more than 7 energy decades from 10 keV to 300 GeV. The GBM extends the energy coverage from below the typical GRB spectral break at ~ 100 keV to above the LAT's low-energy cutoff for inter-instrument calibration. Furthermore, the GBMs sensitivity and FOV will be commensurate with the LAT's to ensure that many bursts will have simultaneous low-energy and high-energy measurements with similar statistical significance. The GBM will also assist the LAT to detect and localize GRBs rapidly by providing prompt notification

of a burst trigger. Finally, the GBM will provide coarse GRB locations over a wide FOV that can be used to repoint the LAT at particularly interesting bursts outside the LAT FOV for gamma-ray afterglow observations, or to notify external follow-up observers. The Principal Investigator of the GBM is Dr. C. A. Meegan (Marshall Space Flight Center) and the Co-P.I. is Dr. G. Lichti from the Max Planck Institute for Extraterrestrial Physics. The GBM collaboration includes scientists from the following organizations:

the Marshall Space Flight Center; the Max Planck Institute for Extraterrestrial Physics (MPE); the University of Alabama in Huntsville, the Universities Space Research Association (USRA), the Los Alamos National Laboratory.

The Marshall, University of Alabama, and USRA scientists are located at the National Space Science and Technology Center in Huntsville, Ala.

4 Overview of *GLAST* Data Transfer

As described in Appendix A of the *Science Data Products Interface Control Document* (Band, D., 2006d), *GLAST* data from the Mission Operations Center or processed *GLAST* data from the Instrument Operations Center will be transferred to the *GSSC* bundled as a Unix tape archive (tar) file. A typical tar file will contain:

1. the data files to be transferred

2. message files accompanying the data (observatory operational command files, not for public access) and

3. a manifest file listing all files/data files and message files (but not the manifest file itself) included in the tape archive file.

Transfers are accomplished using the FASTCopy created by Softlink Ltd. Tar files are named as `SRCID_yyyydddhhmmss.tar` where:

SRCID identifies the origin of the file as either:

MOC Mission Operations Center

GSSC Science Support Center

LISOC LAT Instrument and Science Operations Center

GIOC GBM Instrument Operations Center

yyyydddhhmmss is the timestamp giving the date and time when the tar file was created, en-

coded as a four digit year, a three digit day of year, two digit hour, two digit minute, and two digit second.

As described in Horner, D. (2006), *Ingest System Detailed Design Document* the *GSSC* uses the OPUS pipeline software to un-tar the data and store it at the *GSSC*. The OPUS pipeline will thus populate the *GLAST* data archive. See (Horner, D., 2006) for detailed information, and see §6.7 and Figure 1 for a description of the archive directory structure. Note however that not all data included in the tar file and transferred through the OPUS pipeline will necessarily be publicly available. Restricted transferred data might include spacecraft commands or other sensitive information.

These tar files will be stored by the *GSSC* for a period of two weeks, and will be discarded afterwards.

5 The *GLAST* Archive: Overview

5.1 Archive Components

The *GLAST* archive consists of:

- The observation data archive, i.e. the collection of *GLAST* information describing the events detected by the LAT and/or the GBM, along with relevant information about the performance of the spacecraft and the state of the spacecraft environment as a function of time and/or other relevant parameters (pointing direction, high voltage states, charged particle background, temperature, etc.) during the mission. These data are contained in FITS-formatted files according to the conventions described in Band, D. (2006c), or as catalogs of derived source parameters, or as commonly-used formats (jpegs, gifs, plain ASCII text files) for quicklook summary information and burst alert records;

- The calibration archive, i.e. the collection of data describing the performance of the LAT and GBM instruments in their response to events from in flight and/or ground-based measurements;

- The science tools archive, i.e. the collection of software which can be used to reduce and analyze data from the LAT and the GBM, which includes project-supported software but may also include user-contributed software (provided "as-is");

- Hardware to store and serve these components;

- The interface archive, i.e. the collection of methods which can be used to identify and access *GLAST* science data and *GLAST* analysis tools;

- The *GLAST* Documentation Library, including the *GLAST* Technical Handbook, the Project

Data Management Plan, analysis recipes, etc.; and

- internal *GSSC* databases and configuration files, which serve to document the history of the observatory and the data flow but are usually not available to the general community.

All these products will need to be preserved by the *HEASARC* after the end of GSSC operations.

5.2 The Relation between the *GSSC* and *HEASARC* Data Archives

The relation between the *HEASARC* and *GSSC GLAST* data archives is documented in *The GSSC-HEASARC Memorandum of Understanding*, (MOU; Davis, D. S. and Corcoran, M. F., 2006), which is the controlling document. During the operational life of the *GLAST* mission the *GSSC* has responsibility for providing data to the scientific community, while the *HEASARC* will have the primary responsibility to ensure access to and validity of *GLAST* data after the close of the *GSSC*. As described in the MOU, the *GSSC* will use the *HEASARC* archive infrastructure as the repository of *GLAST* data, which simplifies the handover at the *GSSC* end of operations. The *HEASARC* infrastructure consists of

- hardware (data servers and disk arrays),
- a set of user interfaces (web access methods, like W3Browse, and tools like Skyview, WebPIMMS and other tools),
- data analysis packages and a standard build system,
- protocols for data implementation, access and analysis.

The *GSSC* will receive Level 0 data from the Mission Operations Center (MOC) and Level 1, 2, and 3 data from the Instrument Operations Centers (IOCs). Because of the large volume of LAT data, the *GSSC* maintains a dedicated multi-node computer cluster (a "`beowulf`" cluster) to store versions of individual LAT photon events files optimized for quick searches to provide the scientific community the means to quickly generate FITS-formatted files containing selected photon event data of interesting sources, time intervals or other user-specified criteria. Disks, containing the complete set of *GLAST* LAT photon data in standard FITS formats specially optimized for searching (products LS-001, LS-002, and LS-005, see Table 6.1) are attached to each node of this cluster, so that data retrieval requests can be shared in a load-balanced way. With the concurrence of the *GLAST* project and the *GSSC*, the *HEASARC* may also make these FITS files available for download by `ftp` or some other method. After the close of the *GSSC* the *HEASARC* will ensure that suitably fast access to the LAT data is available to the scientific community either through the continued support of the *GSSC* `beowulf` cluster or by some other method. The guiding principle is that there is to be no loss of capability after the transition from the *GSSC* to the *HEASARC* as the primary server of *GLAST* data.

GLAST observation data (Level 1, Level 2 and Level 3; see §6.1) produced by the *GLAST* project will be transferred to appropriate archive directories by the *GSSC* OPUS pipeline. These data will be transparently accessible from the *GSSC* website and the *HEASARC* through appropriate access methods (W3Browse, ftp, wget, curl, etc). These data include the "optimized" LAT photon and events data (LS-001 & LS-002; see §6.1.2 and Table 6.1). Access to these "optimized" datafiles by the user community will primarily be via the LAT photon data server interface.

Table 1: *GLAST* Archive Components

Component	Size by EOM	Location	Comments
Data Archive	10 TB	GSFC	transfer management from *GSSC* to *HEASARC* at *GSSC* EOM
Calibration	1 TB	GSFC	*HEASARC* CALDB format
Science Tools	10 GB	GSFC	HEASoft compatible tools
Websites	$\sim$ 1 GB	GSFC	*GSSC* website only; sites maintained by other mission elements are the responsibility of those elements
Hardware	Servers, disks	GSFC	Inc. `beowulf` cluster

5.3 Hardware

The physical archive for the *GLAST* mission data, including data from the LAT and the GBM is housed on hardware located at the Goddard Space Flight Center (*GSFC*) in Greenbelt, Maryland. As described in the GSSC-HEASARC memorandum of understanding, during the *GLAST* mission the GSSC is responsible for providing disk space to store data from the LAT and the GBM and for ensuring the completeness and validity of the data on these disks. This storage will be attached to the HEASARC data servers, and will be backed up during normal HEASARC archive backups. An exception to this is the LAT photon, events and Pointing Livetime and Mode History databases, which are composed of sets of FITS files using a "Hierarchical Triangular Mesh" tessellation which is optimized for fast searching. These databases are stored on disks attached to the GSSC `beowulf` cluster. The HEASARC will keep a separate version of these data files primarily as an online backup. After the GSSC close of business, the HEASARC will have complete responsibility for managing all the *GLAST* hardware and will assume responsibility for maintaining the completeness and validity of the *GLAST* data.

5.4 Security

The *GLAST* archive is secured behind the Goddard Space Flight Center (*GSFC*) firewall. This prevents access to data except through http and other common internet protocols. Direct shell access to machines which have the *GLAST* archive disks mounted is available on a restricted basis from machines within the *GSFC* campus. Shell access to these machines from outside the *GSFC* campus is available only via virtual private network (VPN) accounts with special authorization from the *GSFC* Center Network Environment (CNE).

5.5 PGP Encryption

Access to some data may need to be restricted via appropriate methods. Encryption of such data via PGP or other suitable methods may be done. There are no current plans to encrypt any data. Access control to any non-public or proprietary data may also be accomplished using unix permissions and/or firewalls.

6 Data Products Archive

The *GLAST* project has defined the set of data products – see §5 of Band, D. (2006c,d), which is where all data product descriptions in the rest of this section are defined, and which is the controlling document. All data products to be archived will use the FITS file format, and follow the Office of Guest Investigator Programs (OGIP) FITS conventions which are widely used to describe high energy astrophysical data.

6.1 Overview of *GLAST* Data Products

GLAST data is divided by processing state into data levels. Level 0 data, consisting of cleaned satellite telemetry, is mostly unprocessed, or "raw", data. Table 6.1 lists all defined GBM and LAT data products and their archive locations. Higher level data (Level 1, 2 and 3) are produced by the processing pipeline from lower-level data.

Table 2: *GLAST* Archived Data

ICD ID	Product	Description	# files/day	Latency	Size	Level		FFD Tab.
GS-001	CTIME (daily version)	The counts accumulated every 0.256 s in 8 energy channels for each of the 14 detectors.	14	24 hours after receipt of last input data	230 MB (16 MB /file)	1	GIOC Daily Product	5-1
GS-002	CSPEC (daily version)	The counts accumulated every 8.192 s in 1288 energy channels for each of the 14 detectors.	14	24 hours after receipt of last input data	290 MB (20.6MB /file)	1	GIOC Daily Product	5-1
GS-005	GBM gain and energy resolution history	History of the detector gains and energy resolutions; required for calculating DRMs.	14	24 hours after receipt of last input data	42kB (3kB/ file)	1	GIOC Daily Product	5-1
GS-006	GLAST position and attitude history	History of GLASTs position and attitude, required for calculating DRMs	1	24 hours after receipt of last input data	3MB	1	GIOC Daily Product	5-1

Table 2: *GLAST* Archived Data

ICD ID	Product	Description	# files/day	Latency	Size	Level		FFD Tab.
GS-101	CTIME (burst version)	For each detector, the counts accumulated every 0.256 s in 8 energy channels	14	1 day	16MB (1.15 MB /file)	1	GIOC Burst Product	5-2
GS-102	CSPEC (burst version)	For each detector, the counts accumulated every 8.192 s in 128 energy channels	14	1 day	16MB (1.15 MB /file)	1	GIOC Burst Product	5-2
GS-103	GBM TTE	Event data for the burst	14	1 day	40-60MB (3-4.5 MB /file	1	GIOC Burst Product	5-2
GS-104	GBM DRMs	8 and 128 energy channel DRMs for all 14 detectors	28	1 day	6 MB (0.4 MB /file)	1	GIOC Burst Product	5-2
GS-105 (non-burst trigger)	GBM Trigger Catalog Entry (TCAT)	Classification of GBM trigger with some characteristics	1	1 day, updated periodically	20 kB	1	GIOC Burst Product	5-2

Table 2: *GLAST* Archived Data

ICD ID	Product	Description	# files/day	Latency	Size	Level		FFD Tab.
GS-106 (burst trigger)	GBM Burst or Spectral Catalog Entry (BCAT)	Values of the quantities describing the burst (e.g., durations, fluences)	1	1 day, updated periodically	100-200 kB	1	GIOC Burst Product	5-2
GS-107	GBM TRIGDAT	All the GBMs messages downlinked through TDRSS	1	1 day	50-100 kB	1	GIOC Burst Product	5-2
GS-108	GBM Background Files	Backgrounds for spectral fitting	28	1 day	28kB (1kB /file)	1	GIOC Burst Product	5-2
GS-003	Ground-Initiated TTE	Event data accumulation initiated by a ground command	14	Periodically	40-60MB (3-4.5 MB /file	1	GIOC Update	5-3
GS-004	GBM DRM Database Files	DRM files	TBD	Every 6 months		1	GIOC Update	5-3

Table 2: *GLAST* Archived Data

ICD ID	Product	Description	# files/day	Latency	Size	Level		FFD Tab.
GS-007	GBM PHA Look-Up Tables	Tables of the correspondence between CTIME and CSPEC energy channels and the photopeak energy for each detector	4	Every 6 months	4kB (1kB/ file)	1	GIOC Update	5-3
GS-008	GBM DRM Database Compressed Leaf Files	DRMs on a grid of zenith and azimuth angles	TBD	Every 6 months	100GB	1	GIOC Update	5-3
GS-105 (non-burst trigger)	GBM Trigger Catalog Entry	Classification of GBM trigger with some characteristics	1	Updated periodically after initial file	20 kB	2	GIOC Update	5-3
GS-106 (burst trigger)	GBM Burst or Spectral Catalog Entry	Values of the quantities describing the burst (e.g., durations, fluences)	1	Updated periodically	100-200 kB	2	GIOC Update	5-3

Table 2: *GLAST* Archived Data

ICD ID	Product	Description	# files/day	Latency	Size	Level		FFD Tab.
LS-001	LAT Events	Subset of merit n-tuple for subset of the events telemetered to the ground	Per processing run (6-8 per day)	1 day	250 MB	1	LAT products after each run	5-4
LS-002	LAT photons	Selected parameters from the subset of events identified as gamma-ray photons	Per processing run (6-8 per day)	1 day	25 M	1	LAT products after each run	5-4
LS-003	LAT Livetime Cubes	LAT livetime as a function of sky position and off axis angle	Per processing run (6-8 per day)	1 day		1	LAT Updates	5-4
LS-005	LAT Pointing and Livetime History	LAT orientation and mode at 30 s intervals; used to calculate exposures	Per processing run (6-8 per day)	1 day	100 kB	1	LAT Updates	5-4

Table 2: *GLAST* Archived Data

ICD ID	Product	Description	# files/day	Latency	Size	Level		FFD Tab.
LS-008	LAT Point Source Catalog	Table of detected gamma-ray sources with derived information	On update	N/A	10 MB	3	LAT Updates	5-5
LS-009	LAT Burst Catalog	List and characterization of gamma-ray bursts: location, duration, intensity	On update	N/A	TBD	3	LAT Updates	5-5
LS-010	Interstellar Emission Model	Model for diffuse gamma-ray emission from the Milky Way, input for high-level data analysis; will be refined using GLAST data	On update	N/A	40 MB	Ancillary	LAT Updates	5-5
LS-011	LAT Energy Redistribution	Constants for parameterization of the LATs energy redistribution	On update	N/A	12kB (12kB/ file)	1	LAT Updates	5-5
LS-012	LAT Effective Area	Constants for parameterization of the LATs effective area	On update	N/A	120kB (30kB/ file)	1	LAT Updates	5-5
LS-013	LAT PSF	Constants for parameterization of the LATs point spread function	On update	N/A	64kB (17kB/ file)	1	LAT Updates	5-5

Table 2: *GLAST* Archived Data

ICD ID	Product	Description	# files/day	Latency	Size	Level		FFD Tab.
SS-002	Pulsar Ephem.	Ephemerides of pulsars that may be detectable by the LAT	TBD	N/A	TBD		GSSC Product	5-7

6.1.1 Level 0 Data

Level 0 data consist of raw telemetry for the LAT or the GBM which has been subjected to minimal processing to correct for packet duplications or other anomalies. Level 0 data is not suitable for scientific analysis without processing into higher level data. There are no plans to make *GLAST* Level 0 data available to the public at the HEASARC.

Public Access: There is no planned access to the Level 0 data by the general astronomical community.

Archive Plans: Level 0 data from the GBM are transferred to the *GSSC* via the internet from the Missions Operations Center (MOC); Level 0 data from the LAT is also transferred to the *GSSC* from the MOC, after which it is cleaned of ITAR-related telemetry and sent to the LAT IOC (or LIOC). After 6 months (TBR) from receipt of the Level 0 data from the MOC, the Level 0 data is written to SDLT (or possibly, LTO4) digital tape, and then these tapes are sent for storage to the National Space Science Data Center (NSSDC) physically located in building 26 at GSFC.

Frequency: The Level 0 data will be received by the *GSSC* from the MOC daily. These data are then written to tape TBD weeks 6 months (TBR)after receipt of the data from the MOC, and then sent to the NSSDC.

Backup: Data will not be backed up by the GSSC. Prior to EOM, Level 0 may be maintained at the LISOC and GIOC.

6.1.2 Level 1 Data

Level 1 data result from automatic pipeline processing of Level 0 data. In LAT Level 1 processing, the Level 0 data describing the interactions within the LAT will be analyzed to identify and characterize the interacting particle (e.g., photons, electrons, protons, etc.). The Level 1 data for an event will include at least the event arrival time, apparent energy and apparent origin on the sky. Other LAT Level 1 data will include histories of the instrument live time and pointing.

GBM Level 1 processing will primarily re-format and reorganize the GBM Level 0 data. The Level 1 data will consist of continuous (daily) and burst data. Continuous (daily) data are the rates in all GBM detectors in different energy bands, regardless of whether a burst has been detected. Burst data include counts, rates, exposure times, responses, catalog information (e.g., fluence, duration, peak flux), and ancillary data necessary for analyzing the burst.

Public Access: LAT Level 1 data (event lists, photon lists, and pointing history will be publicly available one year after launch (L+1 year). GBM Level 1 data will be available immediately after in-orbit checkout (L+60 day; see §6.3)

Archive Plans: LAT and GBM level 1 data will be transferred by the *GSSC*'s OPUS pipeline

from the IOCs to the *GSSC* disks and to the *HEASARC* data server. Copies of the optimized LAT Level 1 photon and event data will also be stored on the *GSSC* `beowulf` cluster for access by the LAT photon database search engine maintained by the *GSSC*.

Frequency: The frequency of Level 1 data transfer is specified in the Ingest System Detailed Design Document (GSSC-0009)(Horner, D., 2006).

Backup: The copy of the optimized LAT Level 1 photon and event data hosted on the *HEASARC* RAID system serves as an on-line backup of the data. The *HEASARC* will routinely back up these data in the course of normal *HEASARC* backups to an off-site storage facility.

6.1.3 Level 2 Data

Level 2 data results from routine scientific analysis, usually using the science analysis software developed for more focused studies by general scientific community (including GIs) and the instrument teams. For LAT observations these data may include: exploratory science analyses; quick-look analyses to detect transient sources and to support operations planning; standard analysis of transient sources; refined analyses of on-board GRB and AGN transient alerts; and LAT sky maps accumulated over a variety of time intervals. For GBM observations Level 2 data might include the uniform fitting of GRB spectra with standard spectral models.

Public Access: LAT Level 2 data will be publicly available one year after launch (L+1 year). GBM Level 2 data will be available immediately after in-orbit checkout (L+60 days; see §6.3)

Archive Plans: GBM level 2 data will be transferred to the *GSSC* from the GIOC and stored by the OPUS pipeline on disks mounted on the *GSSC* data server and on the *HEASARC* data server. LAT Level 2 data will be transferred from the LISOC to the GSSC and stored by the OPUS pipeline on the GSSC "master disk" (see §6.5) and on the *HEASARC* RAID system.

Frequency: The frequency of Level 1 data transfer is specified in the Ingest System Detailed Design Document (GSSC-0009)(Horner, D., 2006)

Backup: The copy of the Level 2 data hosted on the *HEASARC* RAID system serves as an on-line backup of the data. The *HEASARC* will routinely back up these data in the course of normal *HEASARC* backups.

6.1.4 Level 3 Data

Level 3 data will consist of catalogs and compendia of Level 2 data. The LAT team will produce a catalog of gamma-ray sources, including (but not limited to) flux histories and tentative source identifications. The first LAT catalog will be based on the first-year sky-survey data; updates are to be released following the 2nd and 5th years of operation, and the end of the mission. The GBM team will release catalogs of GBM burst energy spectra. Both instrument teams will maintain

catalogs of transient events. Level 3 data will form the basis of the metadata tables which can be queried via online interfaces (eg. BROWSE).

In addition to these data products, the LAT team will produce, update and make public the diffuse Galactic interstellar and extragalactic emission models used for the analysis resulting in the LAT source catalogs, since the diffuse emission mus t be known to detect point sources. The diffuse Galactic emission is intrinsically interesting because it results from the interaction of cosmic rays with gas and photons in our galaxy. The *GLAST* mission will also produce a table of pulsar ephemerides for use in timing studies.

Public Access: LAT Level 3 data will be publicly available one year after launch (L+1 year). GBM Level 3 data will be available immediately after in-orbit checkout (L+60 days) (see §6.3)

Archive Plans: GBM level 3 data will be transferred to the *GSSC* from the GIOC and stored on disks mounted on the *GSSC* data server and on the *HEASARC* data server. LAT Level 3 data will be transferred from the LISOC to the GSSC and stored on the GSSC "master disk" (see §6.5) and on the *HEASARC* RAID system.

Frequency: The frequency of Level 1 data transfer is specified in the Ingest System Detailed Design Document (GSSC-0009)(Horner, D., 2006).

Backup: The copy of the Level 3 data hosted on the *HEASARC* RAID system serves as an on-line backup of the data. The *HEASARC* will routinely back up these data in the course of normal *HEASARC* backups.

6.2 Expected Data Volume

The data volume for the LAT and GBM for Level 1, 2 and 3 data is expected to be about 300 GBytes per day, or a yearly total of > 1 TB. With an anticipated 10 year mission for *GLAST*, the total volume of the *GLAST* data archive (including photon events, lightcurves, spectra, etc.) is about 10 TB. The GSSC will provide sufficient storage space (hard disks or similar media) to hold this data volume, as described in the MOU.

6.3 Data Access Policies

In the early stage of the *GLAST* mission (roughly lasting 14 months after launch), while data is being calibrated and data processing pipelines finalized, access to some of the *GLAST* data may be restricted to investigators who are directly associated with the instrument teams. Data regarding transient events triggered onboard the spacecraft will be released immediately, and higher level information on a selected set of sources of interest, and untriggered transients, will also be released to the community.

The *GLAST* data release policy is described at `http://glast.gsfc.nasa.gov/ssc/data/policy/`. This policy depends on the mission phase and may initially be restricted to certain categories of investigators as summarized below.

6.3.1 Investigator Categories

For the purposes of *GLAST* data release there are four categories of scientific investigators:

1. The investigators associated with the instruments teams
2. Four Interdisciplinary Scientists (IDSs) chosen to advise the mission and to carry out major data analysis projects in the first years of the mission
3. Guest Investigators (GIs) whose proposals have been accepted
4. All other scientists

In the early part of the mission (Phases 0 and 1, see below) different categories may have different data rights. After the start of mission phase 2, all categories of investigators have the same data rights for Level 1 and higher data products. Typically only category 1 and 2 investigators will have access to raw *GLAST* data (Level 0 data) in any mission phase.

6.3.2 Mission Phases

Post-launch the *GLAST* mission is divided into 3 phases. The type of data which will be released to the public varies with mission phase. The 3 *GLAST* mission phases are:

Phase 0: the first 60 days after launch when the instruments are turned on and calibrated. During this phase the different operational modes will be tested, and bright sources will be observed.

Phase 1: the first year of scientific operations, beginning at the end of Phase 0. The instrument teams continue to calibrate their instruments while conducting a sky survey.

Phase 2: the rest of the mission after the end of Phase 1.

A summary of the duration of the mission phases and the planned data release during each phase is given in Table 6.3.2.

Phase	Start	Duration	Activities	Data Release
0	Launch (L+0)	60 days (L+60 days)	the instrument teams will use the GLAST Level 1 data to calibrate their instruments and refine the processing pipelines. During this phase the different operational modes will be tested, and bright sources will be observed.	GBM: triggered bursts: metadata (direction, fluence estimate and other key information about the burst immediately on discovery); all data (including photons, lightcurves and spectra) for each detector. LAT: all triggered burst metadata (direction, fluence estimate and other key information about the burst immediately on discovery) and data (including photon lists, lightcurves, spectra, exposure cubes and pointing histories); metadata (fluxes and spectral characteristics) for 20-30 LAT sources of interest, and similar metadata for newly-discovered flare sources. Goal: preliminary LAT point source catalog released 6 months after start of phase 1. See `http://glast.gsfc.nasa.gov/ssc/data/policy/LAT_Year_1_Data_Release.html` for more info on LAT data release
1	end of Phase 0 (L+60 days)	1 year (L+14 months)	routine science operations; refine processing/analysis pipelines; calibrations	Same policy as Phase 0 Non-transient data are generally not available to the scientific community. These data will become public a month after Phase 2 begins. The scientific community unaffiliated with the GLAST mission may not affect the mission's observing plan.
2	end of Phase 1 (L+14 monts)	end of mission	routine science operations	LAT photon data first release about 1 month after start of phase 2; subsequent releases of LAT photon data as data received (no proprietary period). GBM data and other LAT data and metadata released as in phase 1.

Source: `http://glast.gsfc.nasa.gov/ssc/data/policy/summary.html`

6.3.3 Transient Source Data Access

The general guidelines for release of transient sources detected by the GBM or LAT are:

Gamma-ray bursts: All data on gamma-ray bursts that trigger either the LAT or GBM will be released. The prompt data release will include direction, fluence estimate and other key information about the burst immediately on discovery. Individual GBM photon data and technical information for their analysis will be released by the GBM team as soon as practical. (This guideline also covers non-gamma ray burst triggers from the GBM, e.g., solar flares). The definition of "practical" will be refined based on data-processing limitations and feedback to the GLAST teams from observatory users such as Guest Investigators and members of multiwavelength collaborations monitoring GLAST sources. If one of the instruments does not trigger on a gamma-ray burst that is detected by the other instrument, information on any detected signal or upper limit will be released as soon as practical.

Sources of Interest: 10-20 Source of Interest (SOI) will be monitored continuously and the fluxes and spectral characteristics will be posted on a publicly accessible web site. Another 10-20 scientifically interesting sources will be added to this list during the Phase 1 survey, including known or newly discovered AGN determined to be of special interest by the TeV and other communities as well as galactic sources of special interest discovered during the survey.

New transients: The community will be notified when a newly discovered source goes above an adjustable flux level of about $2 - 5 \times 10^{-6}$ photons (> 100 MeV) cm^{-2} s^{-1} for the first time. (This corresponds to a peak counting rate of about 100 counts per hour in the LAT.) The flux level is to be adjusted to set the release rate to about 1-2 per week. A high flux derivative observed for a source having a minimum flux should also trigger a release when statistically significant; the derivative, flux and significance thresholds remain to be defined. The release rate for these sources should be included in the 1-2 per week overall release rate, and should be practical for the LISOC to implement. The prompt data release for categories 2 and 3 will include information on source coordinates, flux and other key information such as critical timescales or spectral properties.

6.4 File Name Formats

The GLAST project has decided on the following scheme for naming GLAST data files:

1. Files should have unique, human-readable names; newer versions of a data product should be distinguishable from earlier versions by the file name. The identity of a file may not depend on its position within the directory structure, although a file's name should allow it to be placed into such a system.

2. File names should have no more than 31 characters. The allowed characters are the letters a-z, the numbers 09, and separators '.' and '_'; note that filenames are lower case. (These limitations are for consistency with ISO 9660 Level 2 specifications.)

3. File names should start with '`gl`' and include (in order, as necessary):

   (a) The logical instrument: `g` (GBM), `l` (LAT), `s` (spacecraft);

   (b) Identifier for the data type, such as '`tte`' for time tagged events;

   (c) GBM detectors are identified by '`n`' (NaI) or '`b`' (BGO) followed by a single digit–hexadecimal is used for the 12 NaI detectors.

   (d) Identifier such as burst '`bnyymmddff`', where `yymmdd` signifies the day and `ff` the fraction of day;

   (e) Identifier for the contact number for that day (`c#`), for data products that will be produced once per data downlink;

   (f) Version number, such as `v03`, starting with 00; and

   (g) Three-character format type as file extension, e.g., `.fit` for FITS file.

6.4.1 LAT Automated Science Processing

(see "Automated Science Processing of *GLAST* LAT Data" by J. Chiang et al., HEAD meeting, 2006, from which much of this section is derived.)

Automated Science Processing (ASP) will be performed by the *GLAST* Large Area Telescope (LAT) Instrument Science Operations Center (ISOC) on data from the satellite as soon as the Level 1 data are available in the ground processing pipeline. ASP provides time-critical science analyses of the Level 1 data in order to facilitate fast follow-up and multi-wavelength observations of transient sources. ASP tasks fall into 2 broad categories: GRB detection and follow-up, and detection and monitoring of flaring sources and other "sources of interest" described in the LAT Data Release Plan. The analyses performed in the ASP include refinement of gamma-ray burst (GRB) positions, timing, flux and spectral properties, off-line searches for untriggered GRBs and gamma-ray afterglows, longer time scale monitoring of a standard set of sources (AGNs, X-ray binaries), and searches for previously unknown flaring sources in the LAT band. Table 3 shows the ASP tasks and the deliverable data products.

6.5 Optimized Photon and Event Databases

The *GSSC* will support a "master disk" partition which will contain a complete copy of all *GLAST* science data, including the FITS data, to be served to the community. Updates to the master disk will be done by the *GSSC* by use of FASTCOPY, DTS, OPUS or other protocols (see Horner, D., 2006, for details). To simplify public access, the HEASARC will maintain a copy of the data on the GSSC master disk on the HEASARC network appliance box. For more information about the

Table 3: ASP Data Tasks and Products

Task	Product	Delivery Method to Community	Proposed Latency
1. GRB position refinement using LAT data	GRB position and error, date, time-of-day, and fluences or upper limits in 3 bands	GCN Notice; Web site	15 minute for GCN notice; 1 day for Web page
2. Blind search for GRBs in L1 data	GRB position and error, date, time-of-day, and fluences or upper limits in 3 bands	GCN Notice; Web site	15 minute for GCN notice; 1 day for Web page
3. GRB afterglow detection and analysis	Flux estimates on various time scales, hardness or spectral index estimates; refined position and errors	GCN Circular; Web site	1 hour for GCN notice; 1 day for Web page
4. Monitoring of DRP "Sources-of-Interest"	FITS files containing coordinates, ID, fluxes and plots of flux on 1 day and 1 week intervals in 3 bands; photon index for E > 100 MeV; $1-\sigma$ unc. or $2-\sigma$ upper limits	Web site	weekly
5. Flare detection of new DRP sources	FITS file containing Coordinates, ID, flare onset and duration (if known), plots of light curve in 3 bands	Email notification; Web site	12 hours; Weekly updates

data directory structure in used by the *GSSC* processing pipeline see "Table 9 - Archive Locations for GLAST Data Types" in Horner, D. (2006).

The "master disk" contains the complete LAT event and photon archive, with data "optimized" according to the Hierarchical Triangular Mesh (HTM) tessellation system used for efficient searching of the LAT event data. Although these files are fully compatible with OGIP fits conventions, because of the way the data in them are organized, these optimized event files may be difficult for the general astronomical community to efficiently search and analyze. Thus these files will not be generally available to the scientific community unless by prior agreement with the GSSC or *HEASARC*. Instead users will have access to LAT photon and event data by two main mechanisms:

- Users can extract LAT photons and/or events (and associated ancillary data like livetime cubes and pointing history information) through the GSSC's "modified BROWSE" interface to the HTM data on the `beowulf` cluster. This is the most flexible way to access LAT event or photon data since the user can specify data by various search criteria (time, position, energy, etc).

- The GSSC will also make available versions of the LAT photons and events in standard event file format accumulated over suitable time intervals (for example, accumulations by day, or week, or month, or year - the exact frequency is currently TBD), along with necessary calibration files (livetime cubes and pointing histories). Users can download these files and analyze them using the Science Analysis Environment (SAE).

6.6 Public vs. Proprietary Data

The GSSC master disk partition may contain proprietary (non-public) *GLAST* data. If so, such data will not be available for download from the HEASARC archive. The HEASARC will serve only those *GLAST* data which has entered the public domain. The GSSC and the *GLAST* project, in accordance with NASA, has the ultimate determination on *GLAST* data rights.

6.7 *GLAST* Archive Topology

The archive uniform resource locator (URL) is `http://heasarc.gsfc.nasa.gov/FTP/glast`. The `glast` directory will contain the following subdirectories:

- `doc`, which contains *GLAST* documentation;

- `nra_info`, which contains information regarding *GLAST* NASA research announcements (AOs);

- `data`, containing the actual *GLAST* level 1 & 2 photon events.

- `calib_data`, a link to the *GLAST* calibration database (*CALDB*).

The `/FTP/glast/data` directory shown in Figure 1 will have the same directory structure as the *GSSC* master disk.

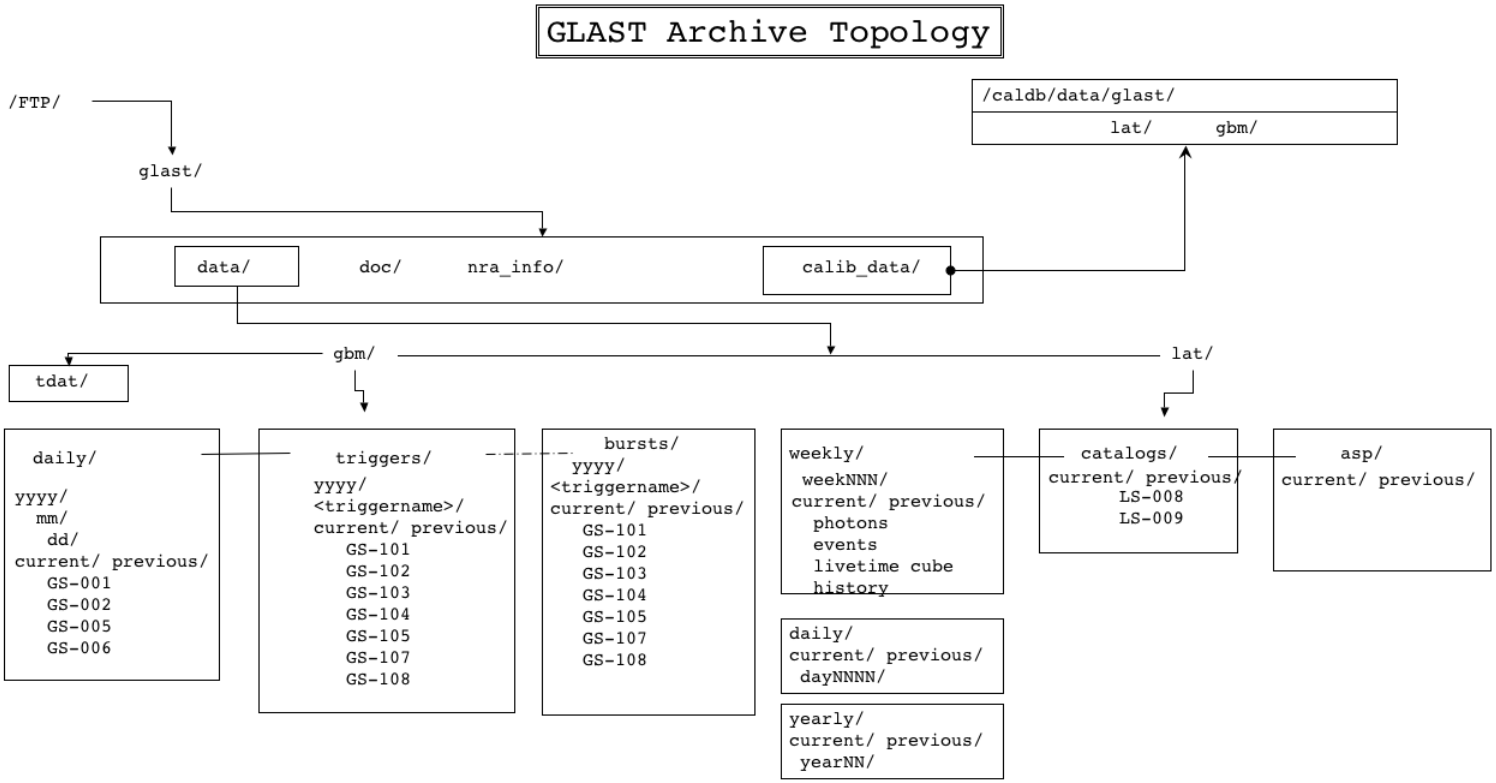

Figure 1: *GLAST* data archive directory structure. The `data/tdat` subdirectory holds the TDAT formatted files used to create tables and catalogs for the HEASARC BROWSE archive interface and is defined and produced by the OPUS processing pipeline. The `data/gbm` and `data/lat` subdirectories hold `FITS`-formatted data files. A dashed line between the GBM trigger data and the GBM burst data is used to show that the burst data are a subset of the triggers, and the burst data directories will be symbolic links to the appropriate trigger directories.

Sample Directory Contents, GBM Daily Data:

/FTP/glast/data/gbm/daily/2009/01/31/

current/

glg_cspec_b0_090131_v01.pha	glg_cspec_n7_090131_v01.pha	glg_ctime_n2_090131_v05.pha	glg_ctime_nb_090131_v05.pha	glg_spechist_n5_090131_v03.fit
glg_cspec_b1_090131_v01.pha	glg_cspec_n8_090131_v01.pha	glg_ctime_n3_090131_v05.pha	glg_poshist_all_090131_v01.fit	glg_spechist_n6_090131_v03.fit
glg_cspec_n0_090131_v01.pha	glg_cspec_n9_090131_v01.pha	glg_ctime_n4_090131_v05.pha	glg_spechist_b0_090131_v03.fit	glg_spechist_n7_090131_v03.fit
glg_cspec_n1_090131_v01.pha	glg_cspec_na_090131_v01.pha	glg_ctime_n5_090131_v05.pha	glg_spechist_b1_090131_v03.fit	glg_spechist_n8_090131_v03.fit
glg_cspec_n2_090131_v01.pha	glg_cspec_nb_090131_v01.pha	glg_ctime_n6_090131_v05.pha	glg_spechist_n0_090131_v03.fit	glg_spechist_n9_090131_v03.fit
glg_cspec_n3_090131_v01.pha	glg_ctime_b0_090131_v05.pha	glg_ctime_n7_090131_v05.pha	glg_spechist_n1_090131_v03.fit	glg_spechist_na_090131_v03.fit
glg_cspec_n4_090131_v01.pha	glg_ctime_b1_090131_v05.pha	glg_ctime_n8_090131_v05.pha	glg_spechist_n2_090131_v03.fit	glg_spechist_nb_090131_v03.fit
glg_cspec_n5_090131_v01.pha	glg_ctime_n0_090131_v05.pha	glg_ctime_n9_090131_v05.pha	glg_spechist_n3_090131_v03.fit	
glg_cspec_n6_090131_v01.pha	glg_ctime_n1_090131_v05.pha	glg_ctime_na_090131_v05.pha	glg_spechist_n4_090131_v03.fit	

previous/

glg_ctime_n0_090131_v01.pha	glg_ctime_n4_090131_v01.pha	glg_ctime_n8_090131_v01.pha	glg_spechist_n0_090131_v01.pha	glg_spechist_n7_090131_v00.pha
glg_ctime_n0_090131_v02.pha	glg_ctime_n4_090131_v02.pha	glg_ctime_n8_090131_v02.pha	glg_spechist_n0_090131_v02.pha	glg_spechist_n7_090131_v01.pha
glg_ctime_n0_090131_v03.pha	glg_ctime_n4_090131_v03.pha	glg_ctime_n8_090131_v03.pha	glg_spechist_n1_090131_v00.pha	glg_spechist_n7_090131_v02.pha
glg_ctime_n0_090131_v04.pha	glg_ctime_n4_090131_v04.pha	glg_ctime_n8_090131_v04.pha	glg_spechist_n1_090131_v01.pha	glg_spechist_n8_090131_v00.pha
glg_ctime_n1_090131_v00.pha	glg_ctime_n5_090131_v00.pha	glg_ctime_n9_090131_v00.pha	glg_spechist_n1_090131_v02.pha	glg_spechist_n8_090131_v01.pha
glg_ctime_n1_090131_v01.pha	glg_ctime_n5_090131_v01.pha	glg_ctime_n9_090131_v01.pha	glg_spechist_n2_090131_v00.pha	glg_spechist_n8_090131_v02.pha
glg_ctime_n1_090131_v02.pha	glg_ctime_n5_090131_v02.pha	glg_ctime_n9_090131_v02.pha	glg_spechist_n2_090131_v01.pha	glg_spechist_n9_090131_v00.pha
glg_ctime_n1_090131_v03.pha	glg_ctime_n5_090131_v03.pha	glg_ctime_n9_090131_v03.pha	glg_spechist_n2_090131_v02.pha	glg_spechist_n9_090131_v01.pha
glg_ctime_n1_090131_v04.pha	glg_ctime_n5_090131_v04.pha	glg_ctime_n9_090131_v04.pha	glg_spechist_n3_090131_v00.pha	glg_spechist_n9_090131_v02.pha
glg_ctime_n2_090131_v00.pha	glg_ctime_n6_090131_v00.pha	glg_ctime_na_090131_v00.pha	glg_spechist_n3_090131_v01.pha	glg_spechist_na_090131_v00.pha
glg_ctime_n2_090131_v01.pha	glg_ctime_n6_090131_v01.pha	glg_ctime_na_090131_v01.pha	glg_spechist_n3_090131_v02.pha	glg_spechist_na_090131_v01.pha
glg_ctime_n2_090131_v02.pha	glg_ctime_n6_090131_v02.pha	glg_ctime_na_090131_v02.pha	glg_spechist_n4_090131_v00.pha	glg_spechist_na_090131_v02.pha
glg_ctime_n2_090131_v03.pha	glg_ctime_n6_090131_v03.pha	glg_ctime_na_090131_v03.pha	glg_spechist_n4_090131_v01.pha	glg_spechist_nb_090131_v00.pha
glg_ctime_n2_090131_v04.pha	glg_ctime_n6_090131_v04.pha	glg_ctime_na_090131_v04.pha	glg_spechist_n4_090131_v02.pha	glg_spechist_nb_090131_v01.pha
glg_ctime_n3_090131_v00.pha	glg_ctime_n7_090131_v00.pha	glg_ctime_nb_090131_v00.pha	glg_spechist_n5_090131_v00.pha	glg_spechist_nb_090131_v02.pha
glg_ctime_n3_090131_v01.pha	glg_ctime_n7_090131_v01.pha	glg_ctime_nb_090131_v01.pha	glg_spechist_n5_090131_v01.pha	
glg_ctime_n3_090131_v02.pha	glg_ctime_n7_090131_v02.pha	glg_ctime_nb_090131_v02.pha	glg_spechist_n5_090131_v02.pha	
glg_ctime_n3_090131_v03.pha	glg_ctime_n7_090131_v03.pha	glg_ctime_nb_090131_v03.pha	glg_spechist_n6_090131_v00.pha	
glg_ctime_n3_090131_v04.pha	glg_ctime_n7_090131_v04.pha	glg_ctime_nb_090131_v04.pha	glg_spechist_n6_090131_v01.pha	

Figure 2: Sample data directory for the GBM daily data. The files include all the data products (for each instrument, where appropriate) as given in Figure 1. Older versions of files are stored online in the `previous` subdirectory.

Table 6.7 gives the directory path for each *GLAST* data type relative to the top-level directory `http://heasarc.gsfc.nasa.gov/FTP/glast/data`, while Figure 1 shows the map of the archive directory structure.

Figure 3: Sample data directory for the GBM trigger data. The files include all the data products (for each instrument, where appropriate) as given in Figure 1. Older versions of files are stored online in the `previous` subdirectory.

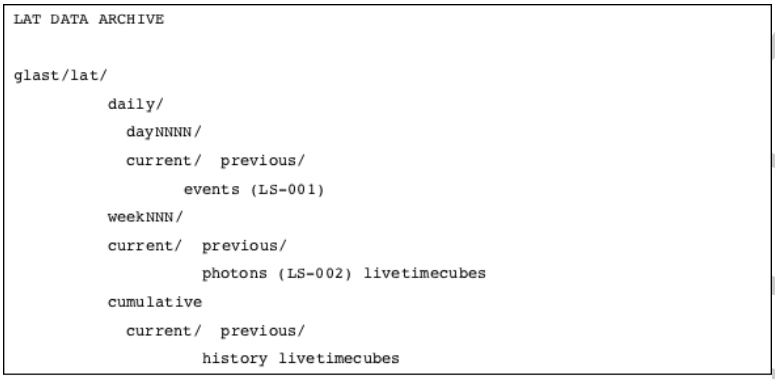

Figure 4: Data directory for the LAT data. Data is accumulated on daily, weekly and yearly timescales. The files include event (product format LS-001), photon (LS-002),pointing and livetime history (LS-003) and livetime cubes (LS-005). Older versions of files are stored online in the `previous` subdirectory.

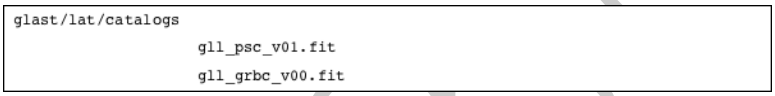

Figure 5: Sample data directory for the LAT catalog data (GRB and Point Source catalogs).

Table 4: *GLAST* Archive Data Directories

ICD ID	Product	Lev.		FFD Tab.	Sample Name	Archive Directory
GS-001	CTIME (daily version)	1	GIOC Daily Product	5-1	`glg_ctime_n0_091125_v00.pha`	gbm/daily/2009/11/25/current
GS-002	CSPEC (daily version)	1	GIOC Daily Product	5-1	`glg_cspec_n0_091102_v02.pha`	gbm/daily/2009/11/25/current
GS-005	GBM gain and energy resolution history	1	GIOC Daily Product	5-1	`glg_spechist_n0_091102_v01.fit`	gbm/daily/2009/11/25/current
GS-006	GLAST position and attitude history	1	GIOC Daily Product	5-1	`glg_poshist_all_091103_v01.fit`	gbm/daily/2009/11/25/current
GS-101	CTIME (burst version)	1	GIOC Burst Product	5-2	`glg_ctime_n1_bn091104_v00.pha`	gbm/triggers/2009/bn091104937/current
GS-102	CSPEC (burst version)	1	GIOC Burst Product	5-2	`glg_cspec_n0_bn091225825_v02.pha`	gbm/triggers/2009/bn091225825/current

Table 4: *GLAST* Archive Data Directories

ICD ID	Product	Lev.		FFD Tab.	Sample Name	Archive Directory
GS-103	GBM TTE	1	GIOC Burst Product	5-2	`glg_tte_n0_bn090131667_v00.fit`	gbm/triggers/2009/bn090131667/current
GS-104	GBM DRMs	1	GIOC Burst Product	5-2	`glg_ctime_n0_bn090202333_v00.rsp`	gbm/triggers/2009/bn090202333/current
GS-105 (non-burst trigger)	GBM Trigger Catalog Entry (TCAT)	1	GIOC Burst Product	5-2	`glg_tcat_all_bn090101001_v00.fit`	gbm/triggers/2009/bn090101001/current
GS-106 (burst trigger)	GBM Burst or Spectral Catalog Entry (BCAT)	1	GIOC Burst Product	5-2	`glg_bcat_all_bn091215020_v05.fit`	
GS-107	GBM TRIGDAT	1	GIOC Burst Product	5-2	`glg_trigdat_all_bn090501102_v00.fit`	gbm/triggers/2009/bn090501102/current
GS-108	GBM Background Files	1	GIOC Burst Product	5-2	`glg_cspec_n1_bn090622050_vxx.bak`	gbm/triggers/2009/bn090622050/current

Table 4: *GLAST* Archive Data Directories

ICD ID	Product	Lev.		FFD Tab.	Sample Name	Archive Directory
GS-003	Ground-Initiated TTE	1	GIOC Update	5-3	`glg_tte_n0_090415500_v00.fit`	
GS-004	GBM DRM Database Files	1	GIOC Update	5-3		
GS-007	GBM PHA Look-Up Tables	1	GIOC Update	5-3	`glg_lutcs_bgo_090319321_v00.fit`	
GS-008	GBM DRM Database Compressed Leaf Files	1	GIOC Update	5-3	`glg_leaf_n0_z125211_az543211_v00.rsp`	
LS-001	LAT Events	1	LAT products after each run	5-4	`gll_ev_r255269026_v000.fit`	
LS-002	LAT photons	1	LAT products after each run	5-4	`gll_ph_r255269026_v000.fit`	

Table 4: *GLAST* Archive Data Directories

ICD ID	Product	Lev.		FFD Tab.	Sample Name	Archive Directory
LS-003	LAT Live-time Cubes	1	LAT Up-dates	5-4	`gll_lt_r255269026_v003.fit`	
LS-005	LAT Point-ing and Live-time History	1	LAT Up-dates	5-4	`gll_pt_r255269026_v001.fit`	
LS-008	LAT Point Source Cata-log	3	LAT Up-dates	5-5	`gll_psc_v01.fit`	lat/catalogs/
LS-009	LAT Burst Cata-log	3	LAT Up-dates	5-5	`gll_grbc_v00.fit`	lat/catalogs
LS-010	Interstellar Emis-sion Model	Ancillary	LAT Up-dates	5-5	`gll_iem_v02.fit`	
LS-011	LAT Energy Redis-tribu-tion	1	LAT Up-dates	5-5	TBD	

Table 4: *GLAST* Archive Data Directories

ICD ID	Product	Lev.		FFD Tab.	Sample Name	Archive Directory
LS-012	LAT Effective Area	1	LAT Updates	5-5	TBD	
LS-013	LAT PSF	1	LAT Updates	5-5	TBD	
	ASP Catalog		GSSC Product		TBD	
SS-002	Pulsar ephem.		GSSC Product	5-7	`gll_psreph_091203_v00.fit` (TBR)	
			GSSC Product		`LAT_allsky_220838401.126_V01.fits`	
			GSSC Product		TBD	
			GSSC Product		TBD	

6.8 Data Transfer from the GSSC to the *HEASARC*

The data on the *GLAST* "master disk", which includes all available GBM and LAT data, will be transferred (via OPUS or rsync or other suitable, reliable protocol) to a dedicated directory under the *HEASARC*'s Network Appliance (NetAPP) RAID unit. This dedicated directory will serve as an on-line backup for the *GLAST* data, and as an alternative data access point for public FTP or http data transfers. *GLAST* metadata tables (see §7) will point to this dedicated area on the NetAPP box when linking to data products. The transfer of data to the NetAPP disk will be accomplished by a suitable mechanism which allows for (at minimum) data transfer logging, checksum verification, and notification of failed transfers.

6.9 Data Reprocessing and Replacement

Both GBM and LAT data may be replaced in the archive due to reprocessing. *GLAST* data files will contain version information as described in §6.4. All versions of files will be stored on the GSSC master disk and on the *HEASARC* mirror. Current versions of *GLAST* data files will reside in the appropriate top-level directory as shown in Figure 1. Older files which have been replaced by newer versions will be moved to "`previous`" subdirectories of the appropriate data directories as shown in Figure 1.

For example, the most up-to-date processing of GBM daily data for 2009 January 31 would be found in the directory
`/FTP/data/gbm/daily/2009/01/31/current`, as shown in Figure 2, while older versions of files belonging to this same dataset would be found in `/FTP/data/gbm/daily/2009/01/31/previous`. The current GBM trigger data for a trigger with the trigger name `bnyymmddff`, where `yy` is the year-2000, `mm` the 2-digit month, `dd` the 2-digit day of month, and `fff` a 3-digit day fraction, for example bn080206937 which occurred in 2008 Feb 6 would be contained in the directory `/FTP/data/gbm/triggers/2008/bn080206937/current`. Previous versions of this dataset would be found in `/FTP/data/gbm/triggers/2008/bn080206937/previous`.

Items of note for GBM `trigger` and `daily` data:

- GBM version numbers start with `00`.
- GBM plans on reprocessing on a file-by-file basis not on a dataset-by-dataset basis. This means that the `currrent` subdirectory may contain version 5 of some files (`v05`) and version 0 of other files (`v00`). It also means that the `previous` subdirectory may not (and probably will not) contain complete sets of files for a given dataset (and a given processing version).
- Not all files will be present for all GBM datasets. There may be datasets for which certain filetypes from some detectors are present but the same filetypes for other detectors are not present. Note that the `current` directories shown in Figures 2 & 3 show all filetypes for all GBM detectors.

6.10 Verification & Validation

The GSSC internal metadata tables maintain the master list of the files, their location in the archive, and the file status, along with information (like md5 checksums or other similar schemes) which allow each file to be verified. The HEASARC will maintain TBD backup versions of these tables, along with MD5 checksum information to allow each metadata table to be verified.

7 Metadata Archive

The metadata tables serve as the community's overview of the *GLAST* archive. These tables contain searchable summaries of *GLAST* sources, *GLAST* pointing and exposure history, and other information. These data tables will be constructed from *GLAST* Level 3 products and ancillary satellite information. Basic tables will use the standard *GLAST* data products from the standard processing of the Level 1 and 2 data from the ISOC for the LAT and GIOC for the GBM. The tables will be constructed by the GSSC and made available for searching via interfaces including the HEASARC BROWSE interface. For additional details see the "Overview of the GLAST Browse Tables" maintained at the *GSSC*, and available at `http://glast.gsfc.nasa.gov/ssc/dev/ingest`.

Tables to be made available via the BROWSE interface will use the HEASARC Transportable Database Aggregate Table (TDAT) format. The summary of *GLAST* metadata tables is given in Table 5.

A specialized interface (a "modified BROWSE" interface, see `http://glast.gsfc.nasa.gov/cgi-bin/ssc/LAT/LEOSimDataQuery.cgi`) will be provided to efficiently search large tables, namely the LAT Event data (LS-001), the *GLAST* pointing and livetime history information (LS-005), and the LAT Photon Summary Data (LS-002) (see §7). These tables may not be available for download in TDAT or ASCII format since the full tables are very large; it is anticipated that these large tables will be stored in individual component files (FITS files fully compliant with the OGIP standards), from which the access software will construct information which satisfies the user's search criteria. Versions of these data accumulated on daily and monthly timescales will be available for direct download by the community.

The detailed structure of the metadata tables and details of their construction is given in the document, "*GLAST* Metadata Tables available at the HEASARC" (in preparation).

While the basic *GLAST* data tables are listed in Table 5, it is anticipated that other metadata tables will be accessible for users via web-based search tools as the *GLAST* mission matures. For example, catalogues of sources resulting from published detailed, focussed analyses of deep fields, or particular source types (AGNs, XRBs) will also be included for online searches as they become available.

Table 5: Metadata Tables for On-line Browsing

Type	Description	BROWSE Shortname	Input Files	Frequency
LAT Point Source Catalogue	Catalog of LAT point sources	GLASTLSRC	LS-008	Periodically
LAT Burst Catalogue	Catalog of LAT bursts	GLASTLBST	LS-009	Periodically
GBM Burst Catalogue	Individual GBM Burst properties	GLASTGBST	GS-106	Periodically
GBM Trigger Catalogue	GBM Trigger properties	GLASTGTRG	GS-105	Periodically
LAT Photon Summary	Parameters of every cosmic gamma-ray photon detected by the LAT			

7.1 Source Catalogs

The GSSC and the HEASARC will make available catalogues of point sources which result from the standard processing by the GIOC and the ISOC will be included as they become publicly available.

The LAT will detect sources and provide a source catalog along with periodic updates. These catalogs in FITS format will be transferred from the LISOC to the GSSC. The GSSC will create TDAT files from these FITS files, and the HEASARC will generate a LAT source catalog from these TDAT files. The goal is to have the source entries available through the BROWSE interface no later than 1 day (TBR) after creation of the TDAT files.

7.2 Burst Catalogs

Both the GBM and LAT will detect bursts and provide burst catalogs. These catalogs in FITS format will be transferred from the IOCs to the GSSC TBD minutes after the burst. The GSSC will create TDAT files from these FITS files, and the HEASARC will generate burst catalogs from these TDAT files. The goal is to have the burst entries available through the BROWSE interface no later than XX minutes (TBD) after creation of the TDAT files, with a requirement of YY hours.

7.3 Trigger Catalogs

Both the GBM and LAT will detect bursts and provide burst catalogs. These catalogs in FITS format will be transferred from the IOCs to the GSSC TBD minutes after the burst. The GSSC will create TDAT files from these FITS files, and the HEASARC will generate burst catalogs from these TDAT files. The goal is to have the burst entries available through the BROWSE interface

no later than 1 day (TBR) after creation of the TDAT files.

7.4 As-Flown Timeline

The As-Flown *GLAST* timeline will be delivered to the *HEASARC* by the GSSC. The timeline will be available from the anonymous ftp area on the *HEASARC* servers in a suitable format (FITS or ASCII, to be reviewed) for download by the user community.

7.5 Verification of Metadata Tables

Metadata made available to users will be verified. Verification will include:

- the HEASARC will directly compare metadata tables in the HEASARC BROWSE system with their source files from which the metadata tables are compiled
- For all table updates, differences between table and updated versions of the table will be determined by the HEASARC and differences noted to ensure that data in common between the original table version and the updated version has not changed.

8 The Calibration Archive

GLAST calibration data is stored at the HEASARC in subdirectories of the *GLAST* CALDB root directory, `/FTP/caldb/data/glast`. The subdirectory `http://heasarc.gsfc.nasa.gov/FTP/glast/calib_data` will be a symbolic link to this directory.

For the GBM, the `pcf` subdirectory holds the "ground-triggered" time tagged events (the `gtte` data) which are GS-003 type files generated by ground command for testing and calibration, along with the "leaf" files generated for a set of off-axis angle ranges for each GBM detector, used for generating the GBM responses. The look-up tables relating channel to energy scale are stored in the `bcf/lut` subdirectories.

For the LAT, the interstellar emission model file and the pulsar database are stored in the `bcf` subdirectory. Responses, effective areas and point spread function files are stored in the `cpf` subdirectory.

Details of the *GLAST* calibration archive can be found in the document "The *GLAST* Calibration Database" (in prep).

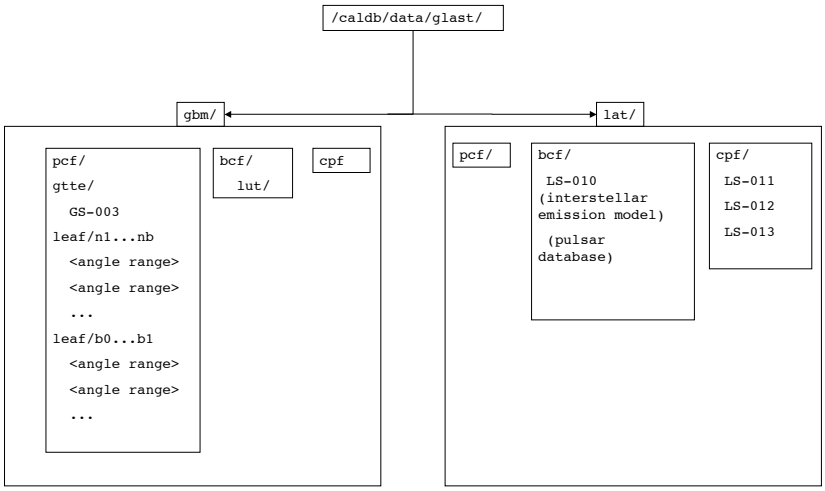

Figure 6: Structure of the *GLAST* Calibration Database (CALDB). Subdirectories `pcf` (primary calibration file), `bcf` (basic calibration file) or `cpf` (calibration product file) hold the calibration data. Versioning is done using standard index files (the `caldb.indx` files).

9 The Hardware Archive

The GSSC hardware (disks, computer systems, including the `beowulf` cluster) will need to be transitioned from control of the *GSSC* to the *HEASARC* after the close of the GSSC. Disks under the management of the GSSC will either be directly mounted onto the *HEASARC* servers; however, incorporating GSSC disks into the *HEASARC* RAID system may incur non-trivial computer administration resources. As an alternative, all data from the GSSC disks may be copied to the *HEASARC* RAID array.

The GSSC `beowulf` cluster is used to create "virtual" photon event lists for *GLAST* LAT data. Management of this cluster may require significant computer administration resources from the *HEASARC*, which may not be available at the time when the GSSC ceases operation. It is important to preserve the functionality provided by the `beowulf` cluster to allow users to extract selected sets of *GLAST* photons. It is unclear before the launch of *GLAST* how best to preserve the functionality of the `beowulf` system after the GSSC closes operation. The *HEASARC* and the GSSC will investigate this issue and come to a mutually agreed-upon plan of action for the resolution of the `beowulf` system at least 2 (TBR) years in advance of the close of the GSSC.

10 Website Archive

The *GLAST* website home page is `http://glast.gsfc.nasa.gov`. There are also individual instrument websites linked from the *GLAST* home page. Table 6 lists the main *GLAST* home pages. All *GLAST* pages hosted at `gsfc.nasa.gov` will be archived at the *HEASARC* after the end of the *GLAST* mission, and access to these pages will be maintained. The maintenance of the individual instrument pages and the Education & Public Outreach pages (which are not hosted at `gsfc.nasa.gov`) will be determined by negotiation between the *GSSC*, the *HEASARC* and the individual organizations.

Table 6: *GLAST* Website Home Pages

Name	URL
GLAST Science Support Center home page	`http://glast.gsfc.nasa.gov/ssc/`
ACD home page	`http://glast.gsfc.nasa.gov/acd/`
GLAST Multi-wavelength Page	`http://glast.gsfc.nasa.gov/science/multi/`
Other *GLAST* websites	
LAT instrument Page	`http://glast.stanford.edu/`
GBM instrument page	`http://gammaray.msfc.nasa.gov/gbm/`
GLAST Education & Public Outreach	`http://glast.sonoma.edu/`

10.1 Integration of the *GSSC*'s Data Access Software and Hardware with the *HEASARC*

Access to public *GLAST* data (for example, pre-packaged *GLAST* source lightcurves and spectra obtained by the LAT &/or GBM; maps of sources or source regions; etc) will be available via the *HEASARC* Browse interface. Some data may be available using a modified interface based on BROWSE and served off of the "beowulf" cluster in order provide faster data access to individual photon events to meet the specific needs of the *GLAST* mission. In the *GLAST* PDMP (Band, D., 2006a), these data are referred to as "optimized databases" in non-archival format, and are accessed using a "modified-Browse" interface. Data products accessed via such a "modified BROWSE" interface are listed in Table 5.1 of the *GLAST* PDMP (Band, D., 2006a), and are reproduced here in Table 7.

Table 7: GLAST Data Available Via a "Modified BROWSE" Interface

Data Product	ICD ID	GSSC Database ID	Frequency
LAT Event Data	LS-001	4110	TBD
Pointing/Livetime History data	LS-005	4120	TBD
Photon Data	LS-002	5110	TBD

During the operational phase of the *GLAST* mission the *HEASARC* will maintain a mirror copy of these "optimized" data products listed in Table 7. These data will not necessarily be made public but will serve as a private, on-line backup in case of problems with the `beowulf` system.

The software which serves the data from the `beowulf` cluster is a set of perl Common Gateway Interface (CGI) scripts which can be run from a single-processor (i.e., non-clustered) environment. If feasible, the *HEASARC* will maintain this access method (i.e. the `beowulf` cluster and associated software) after the end of the *GSSC*. Alternatively it may be that, by the end of the *GLAST* mission, maintenance of this hardware system becomes prohibitive. In this case, prior to the end of the *GLAST* mission, the *HEASARC*, the *GLAST* project scientist and the *GSSC* will provide a mutually agreed-upon mechanism to serve the *GLAST* event and photon data to the user community.

In general, there may be a number of access methods (ftp, specialized web-based queries, etc) available for a given set of data. Any specialized interfaces used by the *GLAST* mission will be available from the *GSSC* website, with links from the *HEASARC* website.

10.2 Web-Based Tools

Table 8 lists web-based tools or applications.

Table 8: *GLAST* Web-based tools

Name	Description	Interface	Data
LAT Event, Photon and Spacecraft Query	`http://glast.gsfc.nasa.gov/cgi-bin/ssc/LAT/LEOSimDataQuery.cgi`	Web Form	LS-001, LS-002, LS-003, LS-005
Gamma-ray background tool	User inputs source or position and region, receives gamma-ray flux in a desired band.	Web Form	Interstellar Emission Model (LS-010)
LAT Exposure Estimatore	User inputs a source or position, receives total amount of LAT exposure at that location	Web form	TBD
GR Weather Map	Animated Display of transient, localized gamma-ray phenomena	Flash animation?	TBD
GR Skymap	All-sky map from GLAST; user can click on a source, get information	clickable map	LS-008, LS-009, GS-106

11 The Software Library

During the mission the *GLAST* Science Support Center (*GSSC*) at *GSFC* is tasked with making *GLAST* data and software analysis tools available to the astrophysical community. The analysis software, the Scientific Analysis Environment (SAE) consists of a set of OGIP-compliant software which can be built using the standard HEASoft build system. The*HEASARC* serves as the ultimate repository and archive for for software from the *GLAST* mission. The goal is to support the SAE as a standard installable HEASoft software package.

11.1 Contributed *GLAST* Software

The *GSSC* may opt to host links to user-contributed software which may be in a non-standard format (perl scripts, idl procedures etc). Any such contributed software will be maintained at the *HEASARC* on a "best-effort" basis. The *HEASARC* makes no guarantee about the utility of any contributed software.

12 Verification Protocols

The GSSC bears sole responsibility for ensuring the scientific validity of the data. The HEASARC is charged with maintaining the *GLAST* archive which includes verification and quality assurance for all public data and metadata. Verification and quality assurance means ensuring that data provided by the GSSC is not changed or corrupted, that all files which are public are publicly available from the HEASARC interfaces, and that all metadata is accurate.

12.1 Final Inventory

Inventory verification means ensuring that all data which are meant to be publicly accessible are in fact publicly accessible. During the operational phase of the GSSC, as the *GLAST* data archive evolves, it is the responsibility of the GSSC to document and ensure that appropriate *GLAST* data is available on the master disk partition. After the end of GSSC operations, when the *GLAST* data archive is static, the GSSC will provide the HEASARC with a full list, known as the "Master Inventory", of the entire master disk partition. For each file on the master disk, the Master Inventory will include:

- the file name;
- the file directory path;
- the MD5 checksum for the file;

- the file size;
- the date of the last update to the file.

The HEASARC will ensure that each file on the Master Inventory exists and is accessible from the HEASARC interfaces and via anonymous ftp or other anonymous internet protocols.

12.2 File-by-File Verification

In addition to ensuring file existence and accessibility, the quality of the individual data files listed on the Master Inventory needs to be verified independently as well. *GLAST* FITS files use the fits DATASUM/CHECKSUM conventions, so that normal FITS verification tools can be used. For each FITS file on the Master Inventory, the HEASARC will run a FITS verification tool like ftverify or a similar verification tool. Any errors revealed by this procedure must be rectified by the GSSC before the *GLAST* archive can be considered final. Such file by file verification will be conducted every TBD weeks while the archive is active, and every TBD months once the archive is finalized.

Non-FITS files must be verified as well. Source metadata catalogs for the LAT and GBM will be verified by a) verification of the FITS-formatted catalog input data used to construct the non-FITS metadata tables, then b) doing a row by row verification of the FITS and non-FITS tables.

13 Handling Reprocessed Data

It is anticipated that *GLAST* data will need to be reprocessed, as new instrumental calibrations are made available, or to correct problems in the processing pipeline which may not be uncovered prior to the population of the data archive. The *GLAST* naming convention (§6.4) naturally allows a distinction between different processing versions. It is not expected however that the version number which appears in the names of a given file will determine for the end-user the reason which caused the data to be reprocessed, or which version of the processing pipeline was used. During the active evolution of the *GLAST* archive, earlier versions of a given file will be maintained in the archive. Once the archive has been finalized, the Master Inventory should contain only those files which are of the highest scientific quality; this means that older versions of files which are no longer deemed to be of the highest scientific quality will not be maintained in the final *GLAST* archive.

14 Information Retrieval

There are a number of ways in which *GLAST* data can be accessed: directly via ftp, through the *GLAST* Photon data browser designed by the GSSC, and via the HEASARC BROWSE interface.

Direct access to the data files and documentation is available via hypertext transfer protocol (http) at `http://heasarc.gsfc.nasa.gov/FTP/glast`

References

Band, D. (2006a). Gamma-ray large area space telescope (glast) project data management plan. Technical report, GSSC.

Band, D. (2006b). Gssc-functional requirements document (rev. a). Technical report, GSSC.

Band, D. (2006c). Science data products file format document. Technical Report GLAST-GS-DOC-0001, GSSC.

Band, D. (2006d). Science data products interface control document (icd). Technical Report GLAST-GS-ICD-0006, GSSC.

Davis, D. S. and Corcoran, M. F. (2006). Memorandum of understanding glast ssc-heasarc (gssc-0008). Technical report, GSSC; HEASARC.

Horner, D. (2006). Ingest system detailed design document. Technical report, GSSC.